



Reader's guide to critical appraisal of cohort studies: 1. Role and design

Paula A Rochon, Jerry H Gurwitz, Kathy Sykora, Muhammad Mamdani, David L Streiner, Susan Garfinkel, Sharon-Lise T Normand and Geoffrey M Anderson

BMJ 2005;330:895-897
doi:10.1136/bmj.330.7496.895

Updated information and services can be found at:
<http://bmj.com/cgi/content/full/330/7496/895>

These include:

References

This article cites 13 articles, 7 of which can be accessed free at:
<http://bmj.com/cgi/content/full/330/7496/895#BIBL>

6 online articles that cite this article can be accessed at:
<http://bmj.com/cgi/content/full/330/7496/895#otherarticles>

Rapid responses

You can respond to this article at:
<http://bmj.com/cgi/eletter-submit/330/7496/895>

Email alerting service

Receive free email alerts when new articles cite this article - sign up in the box at the top left of the article

Topic collections

Articles on similar topics can be found in the following collections

[Other Journalology](#) (385 articles)
[Other Statistics and Research Methods: descriptions](#) (599 articles)

Correction

A correction has been published for this article. The contents of the correction have been appended to the original article in this reprint. The correction is available online at:
<http://bmj.com/cgi/content/full/330/7500/1143-d>

Notes

To order reprints follow the "Request Permissions" link in the navigation box

To subscribe to *BMJ* go to:
<http://resources.bmj.com/bmj/subscribers>

Education and debate

Reader's guide to critical appraisal of cohort studies: 1. Role and design

Paula A Rochon, Jerry H Gurwitz, Kathy Sykora, Muhammad Mamdani, David L Streiner, Susan Garfinkel, Sharon-Lise T Normand, Geoffrey M Anderson

Cohort studies can provide valuable information unavailable from randomised trials, but readers need to be alert to possible flaws

Valid evidence on the benefits and risks of healthcare interventions is essential to rational decision making. Randomised controlled trials are considered the best method for providing evidence on efficacy. However, they face important ethical and logistical constraints and have been criticised for focusing on highly selected populations and outcomes.^{1 2} Some of these problems can be overcome by cohort studies. Cohort studies can be thought of as natural experiments in which outcomes are measured in real world rather than experimental settings. They can evaluate large groups of diverse individuals, follow them for long periods, and provide information on a range of outcomes, including rare adverse events. However, the promise of cohort studies as a useful source of evidence needs to be balanced against concerns about the validity of that evidence.^{3 4}

In this three paper series we will provide an approach to the critical appraisal of cohort studies. This article describes the role and design of cohort studies and explains how selection bias can confound the relation between the intervention and the outcome. The second article will outline strategies for identification and assessment of the potential for confounding, and the third article describes statistical techniques that can be used to deal with confounding. Each paper defines a set of questions that, taken together, can provide readers with a systematic approach to critically assessing evidence from cohort studies.

Randomised trial or cohort study?

Cohort studies are similar to randomised controlled trials in that they compare outcomes in groups that did and did not receive an intervention. The main difference is that allocation of individuals is not by chance. Table 1 gives some important similarities and differences between the two types of study. Because they are expensive and recruiting patients can be difficult, randomised controlled trials are generally short term and used to determine efficacy in selected populations under strict conditions. Cohort studies can be used to determine if the efficacy observed in randomised trials translates into effectiveness in



Cohort studies can use diverse populations

broader populations and more realistic settings and to provide information on adverse events and risks.⁵

Selection bias as a threat to validity

The internal validity of a study is defined as the extent to which the observed difference in outcomes between the two comparison groups can be attributed to the intervention rather than other factors. The biggest advantage of randomised controlled trials compared with cohort studies is that the random allocation process enhances the internal validity of a study by minimising selection bias and confounding.⁶ This paper relies on the definitions provided by CONSORT (box 1).⁷

Allocation by chance in a randomised controlled trial should mean that the groups being compared are similar in terms of both measured and unmeasured baseline factors.⁸ This is not so in cohort studies, and therefore cohort studies are vulnerable to selection bias. In cohort studies, factors that determined whether a person received the intervention could result in the groups differing in factors related to the outcome, either because people were preferentially selected to receive one treatment or because of choices that they made. These baseline differences in prognosis could confound the assessment of the effect of the intervention.

In cohort studies care must be taken to minimise, assess, and deal with selection bias. A comprehensive

This is the first of three articles on appraising cohort studies

Kunin-Lunenfeld Applied Research Unit, Baycrest Centre for Geriatric Care, Toronto, ON, Canada

Paula A Rochon
senior scientist

Meyers Primary Care Institute, Worcester, MA 01605, USA

Jerry H Gurwitz
executive director

Institute for Clinical Evaluative Sciences, Toronto, ON, Canada

Kathy Sykora
senior biostatistician

Muhammad Mamdani
senior scientist
Susan Garfinkel
research coordinator

Department of Psychiatry, University of Toronto, Toronto, ON, Canada

David L Streiner
professor

Department of Health Care Policy, Harvard Medical School, Boston, USA

Sharon-Lise T Normand
professor of health care policy (biostatistics)

continued over

BMJ 2005;330:895-7

Department of Health Policy, Management, and Evaluation, Faculty of Medicine, University of Toronto, Toronto, ON, Canada
 Geoffrey M Anderson
chair in health management strategies

Correspondence to: G M Anderson, Institute for Clinical Evaluative Sciences, 2075 Bayview Avenue, Toronto, ON M4N 3M5, Canada
geoff.anderson@utoronto.ca

Table 1 Comparison of cohort studies and randomised controlled trials

Item	Cohort studies	Randomised controlled trials
Populations studied	Diverse populations of patients who are observed in a range of settings	Highly selected populations recruited on the basis of detailed criteria and treated at selected sites
Allocation to the intervention	Based on decisions made by providers or patients	Based on chance and controlled by investigators
Outcomes	Can be defined after the intervention and can include rare or unexpected events	Primary outcomes are determined before patients are entered into study and are focused on predicted benefits and risks
Follow-up	Many cohort studies rely on existing experience (retrospective studies) and can provide an opportunity for long follow-up	Prospective studies; often have short follow-up because of costs and pressure to produce timely evidence
Analysis	Sophisticated multivariate techniques may be required to deal with confounding	Analysis is straightforward

approach is needed that includes the selection of appropriate comparison groups, the identification and assessment of the comparability of potential confounders between those comparison groups, and the use of sophisticated statistical techniques in the analysis.

Comparison groups in cohort studies

The essence of any cohort study is the comparison of outcomes between people who received the intervention and those who did not. For example, to answer the question, “Do patients who receive an atypical antipsychotic drug have an increased risk of hip fracture?” a cohort study must ask: “What would have happened to these patients if they had not received the atypical antipsychotic drug?”

Ideally, the comparison group in the cohort study should be identical to the intervention group, apart from the fact that they did not receive the intervention. This ideal comparison group is described by methodologists as providing the “counterfactual” or “potential outcome.”⁹ In reality, this ideal comparison group does not exist. Part of the art of designing a cohort study is choosing comparison groups that approach this ideal in order to minimise selection bias while maintaining clinical relevance.

The analysis of the association between antipsychotic drugs and hip fracture can be used to define the types of comparisons that could be found in cohort studies. For any specific intervention (such as exposure to atypical antipsychotics) two factors—the exposure experience of the comparison group and the population from which the intervention and comparison groups are selected—define the types of comparisons that are possible (box 2). People taking atypical antipsychotics can be compared with either people taking an alternative antipsychotic or with those prescribed no antipsychotic drugs. These comparisons could be made in a general population (all elderly people) or in a restricted population (elderly people with dementia).

Questions to ask when assessing a cohort study design

What comparison is being made?

Published studies may include more than one type of comparison, but the focus of any appraisal of a cohort study is on an individual comparison between an intervention group and a comparison group in a defined population. A well written study should contain a clear definition of why the two groups were selected and how they were defined. This information is essential for assessment of clinical relevance and potential for selection bias.

Does the comparison make clinical sense?

The clinical relevance of comparisons needs to be assessed for each case. In the analysis of antipsychotic use and hip fracture, for instance, all four types of comparison might be relevant. However, this might not be true in other analyses. For example, although it would be possible for a cohort study to compare HIV positive patients receiving antiretroviral therapy with those receiving no intervention,¹⁰ this comparison would be irrelevant to many clinicians. A more relevant cohort study would compare patients receiving one antiretroviral therapy with patients receiving another intervention.¹¹ In contrast, a clinically relevant study of the adverse effects of a commonly used treatment such as a non-steroidal anti-inflammatory drug might include a comparison with a no intervention population since no drug treatment could be a realistic option for some people.¹²

Cohort studies should not only describe the populations being compared but also include a discussion of the clinical context for that comparison and provide a justification for the comparison. Readers of these studies should determine if the study makes a comparison that is realistic and relevant to their decision needs.

Box 1: CONSORT definitions of selection bias and confounding⁷

Selection bias—a systematic error in creating intervention groups, causing them to differ with respect to prognosis. The groups differ in measured or unmeasured baseline characteristics because of the way in which participants were selected for the study or assigned to their study groups

Confounding—a situation in which the estimated intervention effect is biased because of some difference between the comparison groups apart from the planned interventions such as baseline characteristics, prognostic factors, or concomitant interventions. For a factor to be a confounder, it must differ between the comparison groups and predict the outcome of interest

Box 2: Possible types of comparisons in cohort study

General population

- 1 Intervention *v* alternative intervention
- 2 Intervention *v* no intervention

Restricted population

- 3 Intervention *v* alternative intervention
- 4 Intervention *v* no intervention

Table 2 Effect on age distribution and sample size of restricting comparison of atypical antipsychotic with no intervention to individuals with dementia

	All older people		Older people with dementia	
	Atypical antipsychotic (n=34 960)	No intervention (n=1 251 435)	Atypical antipsychotic (n=21 427)	No intervention (n=58 754)
Mean (SD) age	80.46 (7.63)	74.50 (6.58)	81.69 (7.11)	80.95 (7.64)
No (%) with dementia	21 427 (61.3)	58 754 (4.7)	21 427 (100)	58 754 (100)

Key questions

What comparison is being made?

Does the comparison make clinical sense?

What are the potential selection biases?

What are the potential selection biases?

Selection bias occurs when there is something inherently different between the groups being compared that could explain differences in the observed outcomes. One powerful strategy to minimise selection bias is to restrict inclusion in the study to those with a defined diagnosis or specific characteristics.³ Restricting the groups to a specific characteristic removes the potential for bias related to that characteristic and can reduce differences in related characteristics. Table 2 presents data from a cohort of older adults given atypical antipsychotics and a no intervention comparison group. Patients taking atypical antipsychotics were over 12 times more likely (63.1% *v* 4.7%) to have dementia. Dementia is related to the risk of hip fracture, and this imbalance may be an important source of confounding. Restricting the study to people with dementia eliminates this source of confounding and reduces selection related to age as the mean age difference between the groups dropped from years to months.

An inevitable consequence of restriction is reduced sample size. In the example, the sample decreased from 1.3 million to about 80 000 when the dementia restriction was applied. When smaller databases are being used, restriction can greatly limit the power of the study. Restriction on the basis of clinical characteristics limits the generalisability of the findings. The more restrictive the population, the less generalisable the results.

It is important to keep in mind the effect the choice of comparison groups will have on potential selection bias when evaluating a cohort study. Some sources of selection bias are clear—for example, if access to atypical antipsychotics was limited to patients of specialists this could result in patients who received these drugs being different from those who did not. Some sources of bias may be more subtle. For example, if doctors thought that atypical antipsychotics had fewer side effects than typical antipsychotics, they might preferentially use the atypical antipsychotics in frailer patients. This form of selection bias, referred to as channelling bias or confounding by indication,¹⁵ occurs when patients are assigned to one intervention or another on the basis of prognostic factors and is key issue in cohort studies.

Readers should recognise the potential for selection bias in all cohort studies and carefully consider possible sources of bias. In the next article we

will outline the link between selection bias and confounding and describe a strategy for identifying and assessing the potential for confounding.

We thank Andreas Laupacis for his comments and Jennifer Gold, Michelle Laxer, and Monica Lee for help in preparing the manuscript.

Contributors and sources: The series is based on discussions that took place at regular meetings of the Canadian Institute for Health Research chronic disease new emerging team. PAR is a geriatrician with extensive research experience in cohort studies of prescription drugs who wrote the first draft of this article and is the guarantor. JHG and MM are clinicians and researchers and SLTN and DLS are statisticians who commented on drafts of this paper. KS programmed and conducted analyses and SG conducted literature searches and reviews. PAR and GMA conceived the idea for the series and GMA worked on drafts of this article and coordinated the development of the series.

Funding: This work was supported by a CIHR operating grant (CIHR No. MOP 53124) and a CIHR chronic disease new emerging team programme (NET-54010).

Competing interests: None declared.

- Gurwitz JH, Col NF, Avorn J. The exclusion of the elderly and women from clinical trials in acute myocardial infarction. *JAMA* 1992;268:1417-22.
- Murray MD, Callahan CM. Improving medication use for older adults: an integrated research agenda. *Ann Intern Med* 2003;139:425-9.
- McKee M, Britton A, Black N, McPherson K, Sanderson C, Bain C. Interpreting the evidence: choosing between randomised and non-randomised studies. *BMJ* 1991;1999:312-5.
- Benson K, Hartz AJ. A comparison of observational studies and randomized, controlled trials. *N Engl J Med* 2000;342:1878-86.
- Black N. Why we need observational studies to evaluate the effectiveness of health care. *BMJ* 1996;312:1215-8.
- Grimes DA, Schulz KF. Bias and causal associations in observational research. *Lancet* 2002;359:248-52.
- Altman DG, Schulz KF, Moher D, Egger M, Davidoff F, Elbourne D, et al. The revised CONSORT statement for reporting randomized trials: explanation and elaboration. *Ann Intern Med* 2001;134:663-94.
- Altman DG, Bland JM. Treatment allocation in controlled trials: why randomize. *BMJ* 1999;318:1209.
- Greenland S, Morgenstern H. Confounding in health research. *Annu Rev Public Health* 2001;22:189-212.
- Badri M, Wilson D, Wood R. Effect of highly active antiretroviral therapy on the incidence of tuberculosis in South Africa: a cohort study. *Lancet* 2002;359:2059-64.
- Fellay J, Boubaker K, Ledergerber B, Bernasconi E, Furrer H, Battegay M. Prevalence of adverse events associated with potent antiretroviral treatment: Swiss HIV cohort study. *Lancet* 2001;358:1322-7.
- Mamdani M, Rochon PA, Juurlink DN, Kopp A, Anderson GM, Nagle G, et al. Observational study of upper gastrointestinal haemorrhage in elderly patients given selective cyclo-oxygenase-2 inhibitors or conventional non-steroidal anti-inflammatory drugs. *BMJ* 2002;325:1-6.
- Psaty BM, Koepsell TD, Lin D, Weiss NS, Siscovick DS, Rosendaal FR, et al. Assessment and control for confounding by indication in observational studies. *J Am Geriatr Soc* 1999;47:749-54.

(Accepted 18 February 2005)

Endpiece

Good advice

Better to hunt in fields, for health unbought,
Than fee the doctor for a nauseous draught.
The wise, for cure, on exercise depend;
God never made his work for man to mend.

John Dryden (1631-1700) in Epistle to John
Driden of Chesterton (1700)

Fred Charatan, retired geriatric physician, Florida

Commentary: Excellent review scheme for critical incidents but insufficient for revalidation

Mayur Lakhani

I want to consider the potential use of the Scottish Audit of Surgical Mortality (SASM)¹ scheme for revalidation of surgeons. Revalidation is an important policy initiative in the United Kingdom from the medical profession's regulatory body, the General Medical Council.² It is aimed at ensuring that doctors remain up to date and fit to practise, and is also a way of restoring and retaining the public's trust in doctors. The policy is in some difficulty, and the government has ordered a review of how revalidation can be made to work.

The Royal College of Surgeons of England and the Senate of Surgery recommend that surgeons include results of clinical outcomes and record of audits (including morbidity and mortality) in their evidence for revalidation.³ To this end, the SASM scheme, which looks at avoidable deaths, seems to be a potentially valuable contribution to the process.

The SASM scheme can be regarded as a peer review of critical incidents. Peer review is an important component of revalidation. The clinical ownership and engagement in the SASM scheme is striking, and there is evidence of the iterative development of standards. There is also clear evidence of improvement resulting from collaboration between clinicians and hospitals.

The disadvantages are that no benchmark is established because the denominator is not known and outliers would not be detected. The analysis concerns itself with the process of surgical care that involved individual surgeons, teams, and the institution, whereas revalidation is an assessment of the individual doctor concerned.

Although patients are involved at a strategic (board) level, lay involvement does not seem to exist at other levels. Surgeons elect members of the management group; this generic procedure (as used by the GMC) was criticised by Dame Jane Smith in her fifth report on the case of the general practitioner Harold Shipman (who was convicted of killing some of his patients and is thought to have killed hundreds more).⁴ Although no evidence exists, this might suggest that the procedure is perceived as a relatively closed process and that it may not meet the modern day requirements of principles of assessment⁵, transparency, and lay involvement.

Revalidation is more than just a record of continuing professional development or taking part in clinical audit. The doctor must also show that his or her clinical performance is not unacceptable—the “patient safety” test. It is significant that participation in SASM is voluntary and that a small number of surgeons do not participate. The reasons for this are not clear, but for the purposes of revalidation a proved and consistent refusal to participate in a national clinical audit scheme focusing on outcomes for surgeons could be a cause for concern.

In conclusion, participation in the critical incident scheme described would be insufficient by itself to revalidate a surgeon. Revalidation should not be its primary purpose. Instead, it is an important and

thoughtful scheme with the potential to develop into a more robust and widespread confidential reporting and learning system to tackle patient safety by focusing on systems improvement.

This is a personal analysis and does not represent a formal view of the Royal College of General Practitioners.

Competing interests: None declared

Royal College of General Practitioners, London SW1 7PU
Mayur Lakhani
chairman of council
MLakhani@rcgp.org.uk

- 1 Thompson AM, Stonebridge PA. Building a framework for trust: critical event analysis of deaths in surgical care. *BMJ* 2005;330:1139-42.
- 2 General Medical Council. *Licensing and revalidation formal guidance for doctors (draft)*. London: GMC, 2004.
- 3 Royal College of Surgeons. *Guidance on surgical practice—criteria, standards and evidence*. 2004. www.rcseng.ac.uk/services/publications/publications/pdf/cse.pdf (accessed 12 Apr 2005).
- 4 *Fifth report. Safeguarding patients: lessons from the past—proposals for the future*. Report of the Judicial Inquiry into Harold Shipman. 2004. www.the-shipman-inquiry.org.uk/home.asp (accessed 12 Apr 2005).
- 5 Postgraduate Medical Training Board. *Principles for an assessment system for postgraduate medical training*. 2004. www.pmetb.org.uk/pmetb/publications/principles.pdf (accessed 12 Apr 2005).

Corrections and clarifications

Necrotising fasciitis

An error occurred in the placement of the three photographs accompanying this Clinical Review article by Saiidy Hasham and colleagues (*BMJ* 2005;330:830-3, 9 Apr), with the result that each is accompanied by the wrong caption. The photograph labelled figure 1 should be accompanied by the caption that was used for figure 2 (“Late signs of necrotising fasciitis . . .”); the photograph labelled figure 2 should be accompanied by the caption that was used for figure 3 (“Split thickness skin grafting . . .”); and the photograph labelled figure 3 should be accompanied by the caption that was used for figure 1 (“Young woman presenting with cellulitis . . .”).

Regulator restricts use of SSRIs in children

This News article by Lynn Eaton (*BMJ* 2005;330:984, 30 Apr) contained two inaccuracies about the drug atomoxetine. The manufacturer, Lilly, points out that we wrongly suggested that atomoxetine is an antidepressant (whereas it has no antidepressant activity). Also, atomoxetine is not associated with an increased incidence of suicide related behaviour (as we implied), although it is associated with an increased risk of hostility (as we stated) and emotional lability.

Interactive case report: Postoperative hypoxia in a woman with Down's syndrome

The table in the first part of this report by A K Siotia and colleagues (*BMJ* 2005;330:834, 9 Apr) was inadvertently given the wrong title during editing. As the text suggests, the table shows the patient's postoperative (not preoperative) results.

Reader's guide to critical appraisal of cohort studies: 1. Role and design

At the final page proof stage of this Education and Debate article, some sort of glitch—the cause of which we have yet to fathom—resulted in the first author's name jumping to the end of the authorship line (*BMJ* 2005;330:895-7, 16 Apr). Paula A Rochon (not Jerry H Gurwitz), therefore, is the first (not the last) author of this paper. This has already been changed on bmj.com.